1-19-2021

Prediction of Alzheimer's Disease-Specific Phospholipase C Gamma-1 SNV by Deep Learning-Based Approach for High-Throughput Screening

Sung Hyun Kim
Neurodegenerative Disease Research Group

Sumin Yang
Neurodegenerative Disease Research Group

Key Hwan Lim
Neurodegenerative Disease Research Group

Euiseng Ko
University of Nevada, Las Vegas

Hyun Jun Jang
Ulsan National Institute of Science and Technology

Correction

NEUROSCIENCE, ENGINEERING

Correction for "Prediction of Alzheimer's disease-specific phospholipase c gamma-1 SNV by deep learning-based approach for high-throughput screening," by Sung-Hyun Kim, Sumin Yang, Key-Hwan Lim, Euiseng Ko, Hyun-Jun Jang, Mingon Kang, Pann-Ghill Suh, and Jae-Yeol Joo, which first published January 4, 2021; 10.1073/pnas.2011250118 (*Proc. Natl. Acad. Sci. U.S.A.* **118**, e2011250118).

The editors note that, due to a printer's error, the article inadvertently published with a number of language errors. The article has been updated online to correct these errors.

Published February 12, 2021.

www.pnas.org/cgi/doi/10.1073/pnas.2101727118

Prediction of Alzheimer's disease-specific phospholipase c gamma-1 SNV by deep learning-based approach for high-throughput screening

Sung-Hyun Kim[a,b,1], Sumin Yang[a,b,1], Key-Hwan Lim[a,b,1], Euiseng Ko[c], Hyun-Jun Jang[d], Mingon Kang[c], Pann-Ghill Suh[b], and Jae-Yeol Joo[a,b,2]

[a]Neurodegenerative Disease Research Group, 41062 Daegu, Republic of Korea; [b]Korea Brain Research Institute, 41062 Daegu, Republic of Korea; [c]Department of Computer Science, University of Nevada, Las Vegas, NV 89154; and [d]School of Life Sciences, Ulsan National Institute of Science and Technology, 44919 Ulsan, Republic of Korea

Edited by Lucio Cocco, University of Bologna, Bologna, Italy, and accepted by Editorial Board Member Solomon H. Snyder December 5, 2020 (received for review June 3, 2020)

Exon splicing triggered by unpredicted genetic mutation can cause translational variations in neurodegenerative disorders. In this study, we discovered Alzheimer's disease (AD)-specific single-nucleotide variants (SNVs) and abnormal exon splicing of the phospholipase c gamma-1 (*PLCγ1*) gene using genome-wide association study (GWAS) and a deep learning-based exon splicing prediction tool. GWAS revealed that the identified single-nucleotide variations were mainly distributed in the H3K27ac-enriched region of the *PLCγ1* gene body during brain development in an AD mouse model. A deep learning analysis, trained with human genome sequences, predicted 14 splicing sites in the human *PLCγ1* gene, one of which completely matched an SNV in exon 27 of the *PLCγ1* gene in an AD mouse model. In particular, the SNV in exon 27 of the *PLCγ1* gene is associated with abnormal splicing during messenger RNA maturation. Taken together, our findings suggest that this approach, which combines in silico and deep learning-based analyses, has potential for identifying the clinical utility of critical SNVs in AD prediction.

Alzheimer's disease | deep learning | *PLCγ1* | single-nucleotide variation

Alternative splicing (AS) occurs in most eukaryotic species and regulates gene expression, giving rise to diverse phenotypes (1, 2). Genetic variants arising due to RNA splicing are more frequently found in individuals having neurodevelopmental disorders, with variable mutation rates between the pre-messenger RNA (pre-mRNA) and mature RNA processing stages (3, 4). These events in genetic mutation reflect marked differences in the balance of transcriptional regulation fidelity among neural networking models (5). A correlation between splicing-mediated mutation and the likelihood of response to neurodegenerative disorders (NDs), together with the identification of associations between gene mutations and clinical outcomes of NDs, can provide comprehensive information on AS for diagnosis. Therefore, detection and evaluation of genetic variations in individuals are crucial for prediction and diagnosis of NDs (6).

Alzheimer's disease (AD) is an ND affecting different brain regions, ranging from the cerebral cortex to the hippocampus (7). Clinically, abnormalities, such as amyloid plaque (known as amyloid beta [Aβ] aggregation) and neurofibrillary tangle, are usually first observed in brain tissues of patients with AD. The progression of these abnormalities to other areas of the cortex is gradual and shows considerable differences in the rate of occurrence among individuals (8). Although biochemical research has contributed to important advancements in the diagnosis of AD pathogenesis, studies on single-nucleotide variants (SNVs) in AD are scarce due to the lack of informative genetic information (9). Therefore, identification of gene mutations in AD remains challenging, and a comparative genomics approach is required.

The products of *PLC* genes are activated by extracellular signaling factors, such as neurotransmitters and hormones, that trigger intracellular signaling receptors such as G protein-coupled receptors and receptor tyrosine kinases (RTKs) (10). Phospholipase c gamma-1 (*PLCγ1*), which encodes a signal transducer of RTKs, has been suggested as a candidate for neuronal development (11). Mutation in *PLCγ2*, a member of the *PLC* gene family, has been evaluated in diverse diseases, such as AD, myelodysplastic syndromes, and chronic lymphocytic leukemia (12–14). However, the analysis of mutations in *PLCγ1* has been limited to T cell lymphomas and angiosarcoma (15, 16). Although *PLCγ1* is known to be associated with AD pathogenesis (17–19), SNVs arising due to posttranscriptional splicing or frameshift remain unknown.

High-throughput screening-based deep learning has been used to predict splicing from primary sequences (4). However, high-throughput screening-based deep learning has not yet been extended to identify SNVs in specific target genes associated with AD. Therefore, we reasoned that genome-wide association study (GWAS)-derived deep learning would be an effective model system for predicting AD and its progression. We used a GWAS platform to analyze query expression patterns and mutations of

Significance

DNA mutation within gene bodies contributes to abnormal translation and can lead to neurodegenerative disorders. High-throughput analysis is suitable for initial detection of gene mutations with details. Deep learning-based RNA splicing analysis facilitates accurate and precise predictions of genetic variants in DNA bodies. Although deep learning-based prediction methods have improved the screening of genetic variations for target diseases, there have been no reports showing a direct comparison with genetic information of an AD model. This study identified the gene mutations and abnormal splicing of *PLCγ1* gene in AD using both high-throughput screening data and a deep learning-based prediction. Our findings provide insight for improvement in prediction and diagnosis of AD pathology.

Author contributions: P.-G.S. and J.-Y.J. designed research; S.-H.K., S.Y., K.-H.L., H.-J.J., and J.-Y.J. performed research; E.K., M.K., P.-G.S., and J.-Y.J. contributed new reagents/analytic tools; K.-H.L., M.K., and J.-Y.J. analyzed data; and S.-H.K., S.Y., K.-H.L., M.K., P.-G.S., and J.-Y.J. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. L.C. is a guest editor invited by the Editorial Board.

[1]S.-H.K., S.Y., and K.-H.L. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: joojy@kbri.re.kr.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2011250118/-/DCSupplemental.

Published February 12, 2021.

the *PLCγ1* gene and found a gradual increase and accumulation of acetylation in the gene during brain development. In addition, we found that exon splicing of the *PLCγ1* gene in mature mRNA was caused by single-nucleotide insertion in AD. The regions of *PLCγ1* gene mutations are evolutionally conserved across various species, including humans. The findings of this study provide insight regarding a previously uncharacterized AD-related gene that may have clinical significance for AD prediction.

Results

Genome-wide Analysis of Gene Expression in Cortex Using High-Throughput Total RNA Sequencing. To examine the transcriptional regulation of AD-specific gene expression, we performed high-throughput total RNA sequencing (RNA-seq) analysis of cortices from 6-mo-old wild-type (WT) and 5xFAD mice. Since the expression levels of many genes are programmed and regulated in tissues and blood of AD mouse models (20–22), genes with increased or decreased expression levels in AD might be functionally important in the brain. We isolated the cortex region from mouse whole brain and extracted RNA for total RNA-seq (Fig. 1*A*). Data from the cortices of WT and 5xFAD mice were obtained to identify differentially expressed genes.

We confirmed that the expression levels of 17,109 genes were 5xFAD specific (Fig. 1*B*). Overall, 1,472 genes were significantly up-regulated, whereas 653 genes were down-regulated in the 5xFAD mouse cortex (Fig. 1*C* and *SI Appendix*, Fig. S1). Previous studies have reported that dysregulations of *PLC* genes are closely associated with many brain disorders, such as schizophrenia, bipolar disorder, Huntington's disease, depression, and AD (10). Although the importance of the correlation between *PLC* genes and AD has been suggested, supporting evidence, such as genes exhibiting differential expression or variation between healthy individuals and those with AD, has not been identified. To demonstrate the expression levels of *PLCβ1*, *PLCβ2*, *PLCβ3*, *PLCβ4*, and *PLCγ1* in the brains of WT and 5xFAD mice, we performed GWAS for *PLCβ1*, *PLCβ2*, *PLCβ3*, *PLCβ4*, and *PLCγ1* using high-throughput total RNA-seq data to measure their transcription levels. The expression levels of *PLCβ1*, *PLCβ2*, *PLCβ3*, *PLCβ4*, and *PLCγ1* were slightly up-regulated in the 5xFAD cortex, while *PLCβ2* expression was up-regulated by ∼2.4 times in 5xFAD (Fig. 1*D* and *SI Appendix*, Fig. S2*B*). However, the endogenous expression level of *PLCβ2* was significantly lower than those of *PLCβ1*, *PLCβ3*, *PLCβ4*, and *PLCγ1* in the 5xFAD cortex (*SI Appendix*, Fig. S2). In addition to the analysis of *PLCβ1*, *PLCβ2*, *PLCβ3*, *PLCβ4*, and *PLCγ1* transcription levels, Western blot analysis was performed to confirm changes at the translational level, revealing that there were no significant differences among the levels of the proteins encoded by these genes in WT and 5xFAD cortices (*SI Appendix*, Fig. S3). Notably, PLCβ2 protein was not detected in the Western blot analysis because its endogenous level in neurons was very low (23).

Single-nucleotide polymorphisms (SNPs) in *APOE* or *TREM2* have been proposed to play pathogenic roles in AD. Moreover, somatic mutations in the brain have been suggested to be associated with genetic architecture in AD (24–26). RNA-seq data are benefitted by the detection of unidentified alternative variations and specific gene expression (27). Though the expression levels of PLC subfamily proteins showed no significant differences between healthy individuals and those with AD, the evidence led us to analyze the genomic variations of *PLC* genes in AD. To obtain the profiles of *PLC* subfamily genes based on SNV data from total RNA-seq, we proposed a pipeline for RNA-seq–based SNV and insertion/deletion (InDel) analyses (Fig. 2). We performed gene expression and SNV profiling of cortex from brains of WT and AD mice using total RNA-seq to obtain unknown information on SNVs in *PLCγ1*, *PLCβ1*, *PLCβ2*, *PLCβ3*, and *PLCβ4*, and provide insight into genetic variation in the pathogenesis of AD.

Identification and Characterization of AD-Specific Mouse *PLCγ1* and *PLCβ* SNVs in the Cortex. High-throughput total RNA-seq is one of the most potent techniques for determining SNVs and genetic variants of target genes. A total of 163 variations (132 SNVs and 31 InDels) were identified in five genes. Of these, most were located in introns (152 variants), and only five variations were located in exons; the remaining six variants were located in splicing regions. In the 5xFAD cortex, *PLCγ1* had 13 variants, while *PLCβ1*, *PLCβ2*, *PLCβ3*, and *PLCβ4* had 66, 6, 3, and 75 variants, respectively (*SI Appendix*, Tables S1–S5). In this study, we mainly focused on variations in the exons of *PLCγ1*. There were five functional variations in the exons, including three synonymous SNVs, one frameshift insertion in exon 27 of *PLCγ1*, and one stop gain in exon 11 of *PLCβ3* (*SI Appendix*, Tables S1 and S4). A frameshift insertion, wherein a single "A" nucleotide was inserted at position 160,759,682 in chromosome 2, gave rise to an abnormal codon that substituted the isoleucine at the 970th amino acid position to asparagine (Fig. 3*A*) in the PLCγ1 protein of the AD mouse model. Stop-gain mutation occurred due to the substitution of the reference "G" nucleotide with "A" at position 6,963,413 in chromosome 2, encoding a terminal codon due to a change of glutamine at the 352nd amino acid of *PLCβ3* in the 5xFAD cortex (*SI Appendix*, Table S4).

However, we found no differences in the translation and termination of PLCγ1 and PLCβ3 between WT and the 5xFAD cortex (*SI Appendix*, Fig. S3). Eukaryotic AS events are closely related to adult tissue functions, organ development, and tissue homeostasis (28, 29); alterations in any of these can affect cell proliferation, methylation, and migration of cancer cells. Aberrant splicing is implicated as an important factor contributing to diseases (30). To determine whether exon 27 of *PLCγ1* affects AS after mRNA processing, we performed RT-PCR using various primer sets targeting the region near this exon. Exons 26 to 30 of *PLCγ1* were deleted in the mature mRNA of 5xFAD cortex. Conversely, exons 27 to 32 of *PLCγ1* were conserved in the pre-mRNA of both WT and 5xFAD cortex (*SI Appendix*, Fig. S4), suggesting that exon skipping or AS occurs at the *PLCγ1* SNV region in AD. Single amino acid substitutions have been linked to many human diseases; these disease-causing mutations tend to cause changes in hydrogen-bonded linkages or bridges, resulting in harmful amino acid mutations (31).

The frequency of AS differs among species. For example, although both human and mouse genomes have similar numbers of conserved motifs, the average alternative pre-mRNA splicing rates of human (>95 to 100%) and mouse (∼63%) genes are different (32). This comparison between human and mouse genes has also revealed that alternative exons of the reading frame are conserved in human and mouse, indicating that genetic changes are caused by AS (32). We further examined whether the 970th amino acid, isoleucine, at exon 27 of mouse *PLCγ1* was conserved in other species. We obtained the *PLCγ1* sequences of various species using the UCSC genome browser. Sequence alignment of *PLCγ1* from various species, including humans, revealed that exon 27 of mouse *PLCγ1* was completely conserved across species. Moreover, exon 26 of human *PLCγ1* was well conserved with respect to the isoleucine site. Surprisingly, the full-length amino acid sequences encoded by exon 26 of human and exon 27 of mouse *PLCγ1* matched completely (Fig. 3*B*). Collectively, our findings suggest that the SNV in exon 27 of mouse *PLCγ1* plays an important role in AS during mRNA processing, and may have potential as a biomarker for prediction of AD.

High-Resolution Profiling of Histone Modification at *PLCγ1* in the Forebrain during Brain Development. Histone acetylation facilitates transcriptional activity and plays an important role in the development of human brain diseases through epigenetic modification events, such as altering chromatin structure or accessibility of transcription factors (33, 34). To determine whether epigenetic

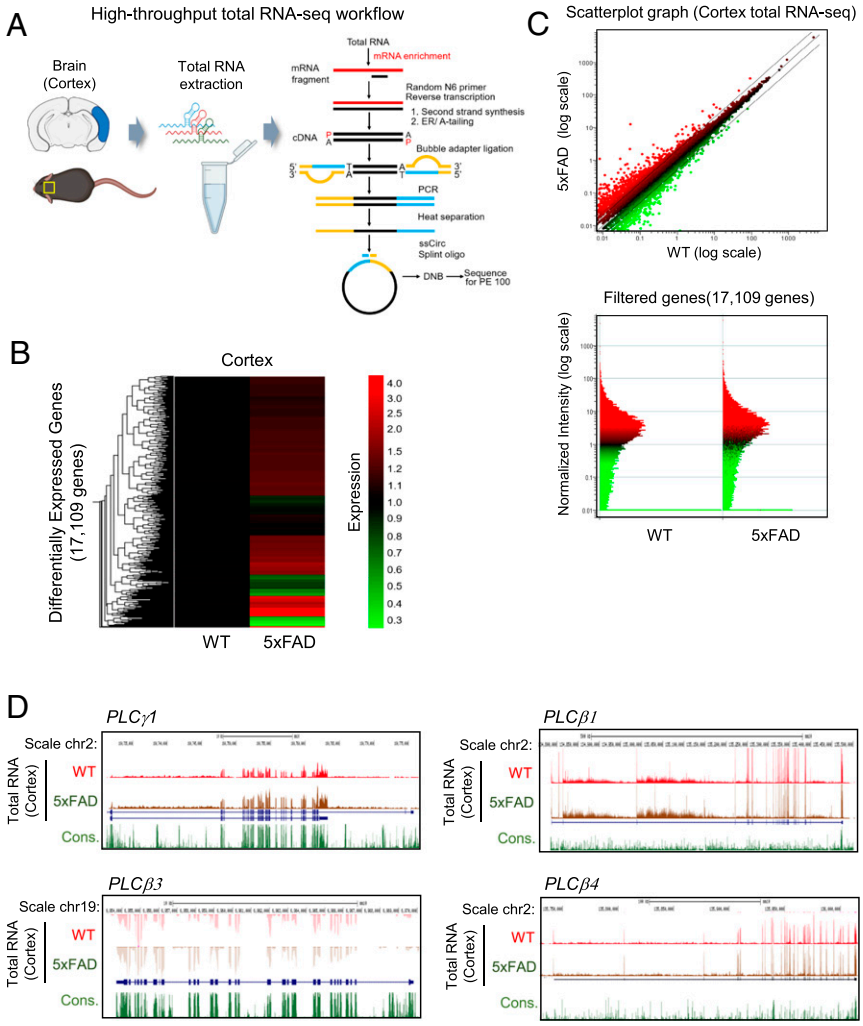


Fig. 1. High-throughput total RNA-seq profile for *PLCγ1* and *PLCβ* subfamily gene expression in the cortex. (*A*) High-throughput total RNA-seq analysis workflow. (*B*) Heat map and hierarchical clustering of total RNA-seq analysis from mouse cortex. The 17,109 genes are differentially expressed between WT and 5xFAD. (*C*) Normalization of total RNA-seq using scatterplot matrices from WT and 5xFAD model mouse cortex. The *x* axis indicates gene expression of WT; the *y* axis indicates gene expression of AD. (*D*) Genome browser view of the *PLCγ1*, *PLCβ1*, *PLCβ3*, and *PLCβ4* genomic loci and expression display with total RNA-seq data in cortex. ER, End Repair; ssCirc, single-strand Circle; DNB, DNA NanoBall; PE 100, Paired end 100; Cons, Consensus.

changes may occur in the *PLCγ1* gene during mouse brain development, we performed H3K27ac enrichment profiling using a chromatin immunoprecipitation sequencing (ChIP-seq) dataset to generate high-resolution histone modifications. Histone acetylation gradually accumulated in the *PLCγ1* gene in mouse forebrain regions ranging from E11.5 to P56. Histone acetylation increased in the introns of *PLCγ1*, and AD-associated SNVs were also concentrated in this region (Fig. 4). In addition, AD-specific nucleotide alternation sites were distributed in the noncoding region of *PLCγ1*, and we found several transcription factor binding motifs at AD-specific AS sites (*SI Appendix*, Fig. S5). These results suggest that SNVs in *PLCγ1* are concentrated in the introns in AD, indicating that they may interfere with the epigenetic role associated with AD.

Prediction of AD-Specific Nucleotide Alteration Sites in the Human Genome Using Deep Learning. Dysregulation of AS has been implicated in AD (35, 36). Recently, a new deep learning-based tool called SpliceAI, which can predict splice junctions of target genes from pre-mRNA nucleotide sequences with high accuracy, was introduced (4). Total RNA-seq–based SNV analysis confirmed the profile of AD-specific SNVs in *PLCγ1* in the 5xFAD mouse cortex. Then, we further examined whether a single-nucleotide alteration in human *PLCγ1* may be correlated with the 5xFAD cortex. After SpliceAI analysis predicted a total of 14 splicing sites in the human *PLCγ1* gene, accurate delta scores and positions were analyzed (Fig. 5), and a novel splicing site in exon 26 of human *PLCγ1* was identified. The single nucleotide "G" at position 41,172,421 in chromosome 20 was substituted with "A," "C," or "T" at position 41,172,423, which was completely correlated with exon 27 of mouse *PLCγ1* in the AD model cortex. SpliceAI analysis accurately revealed that the SNVs in exon 26 of human *PLCγ1* and exon 27 of mouse *PLCγ1* were correlated. These AD-associated SNVs occurred at the same position in humans and the 5xFAD cortex. Taken together, these results clearly demonstrate the prediction accuracy of SpliceAI in humans with respect to genetic mutations and splicing in the AD model.

Discussion

Given the fact that genetic variants exist in mammals at different stages of development, it is likely that these variants play roles in determining phenotypes and adaptation to environments as response to unexpected stimuli (37). It frequently has been questioned whether genomic information is directly associated with

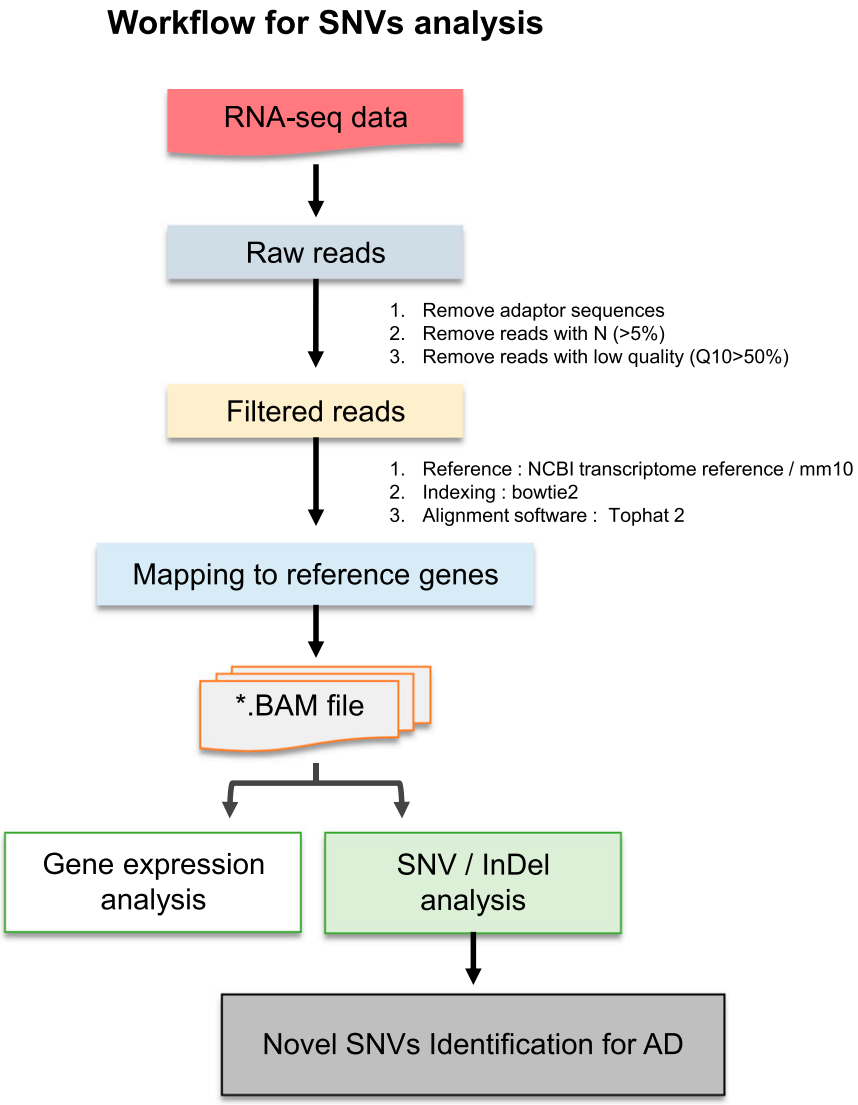


Fig. 2. Total RNA-seq–based SNV identification workflow for AD. Single and paired-end reads were calculated using next-generation sequencing. Mapping was performed with reference genes and each bam file was indexed with SAMtools. Preprocessing formed Ras sequencing data. SAMtools mpileup was performed with University of California Santa Cruz (UCSC) mm10 genome as a reference, and the SNV/InDel was then analyzed for AD. VCF files were generated with vcfutils.pl varFilter, and functional annotation of each variant was performed with ANNOVAR (ANNOtate VARiation) software.

the expression to determine phenotypes. In addition, a comparison of expression at transcriptional and translational levels in the same tissue may not be appropriate (38). Although gene copy number variations and gene copy number alterations (CNAs), which usually lead to changes in mRNA levels, are believed to contribute toward the development of several diseases such as tumors, CNAs leading to transcriptional changes do not cause translational changes (39). A previous study evaluated genetic variation leading to changes at the transcriptional level and in ribosome profile using RNA-seq and GWAS (40). However, correlations between genomic variation and protein expression in NDs, such as AD, are not fully understood yet. Therefore, we performed careful measured mature RNA and analyzed splicing patterns in genetic variants using GWAS-based deep learning to provide insight into individual SNVs in AD.

This study has at least two major implications. First, genetic variants arising due to gene mutations and RNA splicing are respectively conserved and observed in AD. Second, these variants could be predicted using deep learning-based tools to identify therapeutic and diagnostic targets for AD. The GWAS of AD brain tissues described here was instrumental in defining genetic variations. This study evaluates a deep learning-based analysis for the detection of gene variants in AD and examines the phenotypes of these variants and their association with disease progression. The marked increase in the level of histone acetylation in introns during development presumably enables highly specific activation of long noncoding RNAs as enhancer RNA, which may be needed to mediate the expression of some proteins (41, 42). H3K27ac is a well-known marker of active enhancers and is associated with enhancer activity (43, 44). Superenhancers, which are large clusters of enhancers, play an important role in biological regulation (transcription and translation) and various human diseases. Indeed, many diseases are associated with variations in superenhancer-enriched regions (45). Previous studies have identified SNPs associated with various human diseases, such as AD, type 1 diabetes, white blood cell distribution, fasting insulin level, and so on (45). For example, two SNPs in the superenhancer region of BIN1 gene are associated with AD (46). Moreover, trait-associated SNPs occur in noncoding regions within enhancers, where H3K27ac accumulates (45).

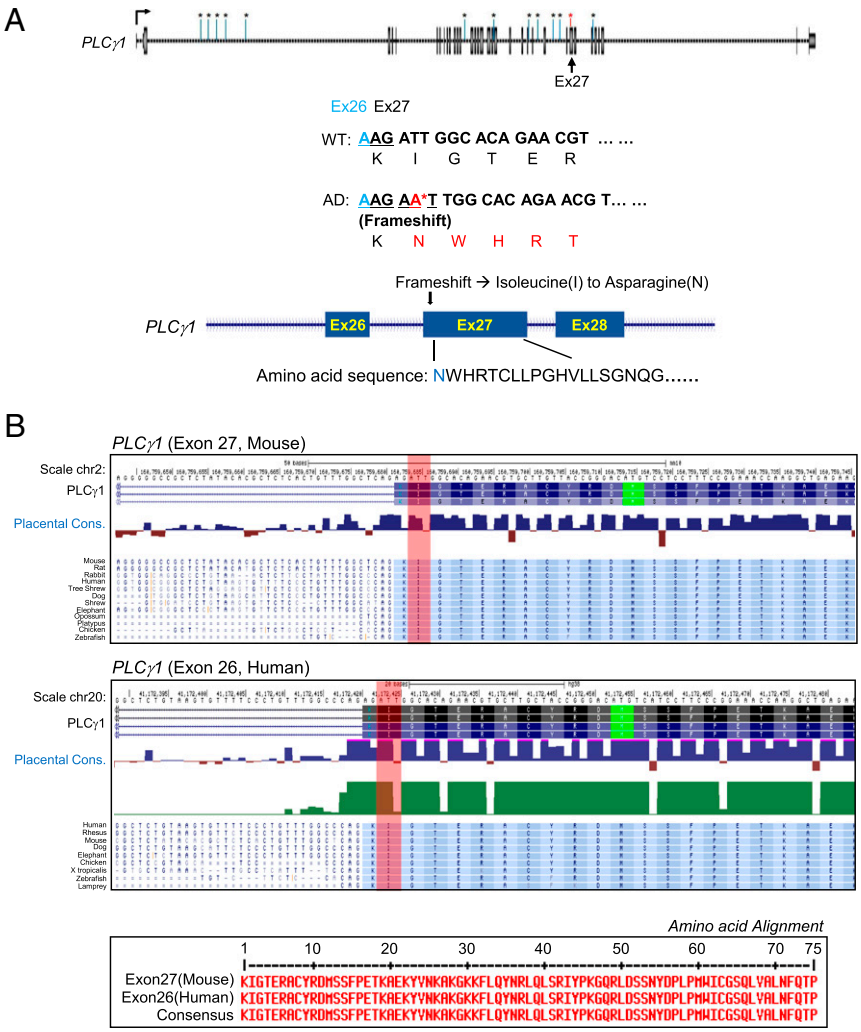


Fig. 3. Identification and characterization of AD-specific sequenced mouse *PLCγ1*. (*A*) Mapping for AD model mouse-dependent frameshift at exon 27 of the *PLCγ1* gene body. One single-nucleotide insertion was caused by an abnormal codon by breaking the 970th amino acid, isoleucine, into asparagine. (*B*) UCSC genome browser sequence alignment of various species, including the human *PLCγ1* gene. Both exon 27 of mouse *PLCγ1* and exon 26 of human *PLCγ1* were well conserved in isoleucine site, and the full length of the amino acid sequence was completely matched. Red vertical bars indicate the locations of the isoleucine for both human and mouse *PLCγ1*.

AD-specific nucleotide alternation sites are distributed in non-coding regions of the *PLCγ1* gene, and we identified several transcription factor binding motifs in AD-specific alternative sites (*SI Appendix*, Fig. S5). These variations can also be attributed to RNA maturation and splicing (Fig. 4). Consistent with these findings, epigenetic changes of target genes are also correlated with AD.

Among the unresolved questions regarding molecular mechanisms of genetic variations, including RNA splicing in AD progression, the potential of deep learning for application in prediction of diagnostic targets for diseases remains obscure. For example, although the genetic risk factors associated with AD progression mainly support the importance of Aβ aggregation, it has been indicated that Aβ is necessary but not sufficient for AD progression, and other unknown factors may play a key role (47). However, identifying such unknown factors is difficult without specific genetic analysis. Therefore, deep learning, including genome footprinting, has now become an alternative method to predict diseases (48). Through genetic analysis and exon mapping, we found that mature RNA was modified through splicing in an AD model (*SI Appendix*, Fig. S4), suggesting that genetic mutations might be conserved in humans and that they can be analyzed with deep learning of pre-mRNA nucleotide sequences. Surprisingly, predictions of genetic mutations in human *PLCγ1* matched exactly with the genome variant data of the AD model (Figs. 3 and 5). Although a wide range of clinical phenotypes is observed in AD patients, a liquid biopsy of bone marrow, blood, and urine is required for the diagnosis of AD. We also analyzed the SNVs in *PLC* genes and their expression in the blood of the AD model; the result indicated that AD-specific novel *PLCβ1*, *PLCβ2*, *PLCβ4*, and *PLCγ1* SNVs in the blood are different within the cortex (*SI Appendix*, Table S6). *PLCγ1* expression was enriched in AD patients, according to peripheral blood mononuclear cells microarray data, and this was consistent with the results of GWAS analysis in the AD mouse model (*SI Appendix*, Fig. S6). Collectively, GWAS-based deep learning results may be applied to predict the clinical outcome of AD patients and could contribute to both diagnosis and clinical treatment of AD.

In conclusion, we have demonstrated the importance of GWAS-based deep learning analysis for predicting genetic variation in AD. This study identifies the SNVs in the AD model through a combination of high-throughput screening and deep learning analysis. Given the DNA sequence of the target gene, our method enables the prediction of transcriptional events based on RNA splicing

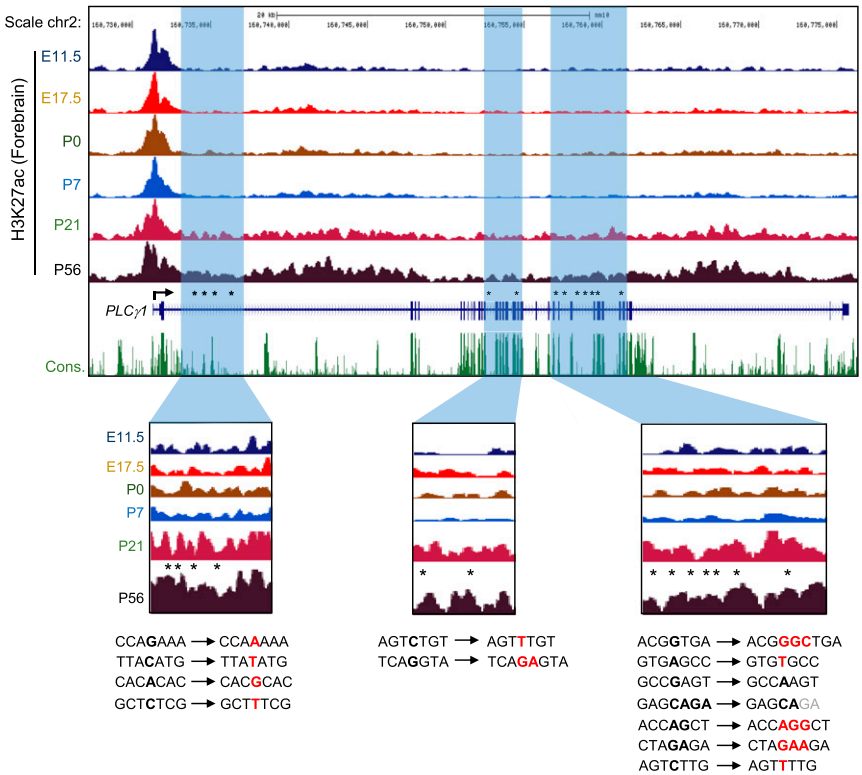


Fig. 4. High-resolution H3K27ac binding profiles of histone acetylation in mouse forebrain during brain development. Histone acetylation was highly accumulated in the *PLCγ1* gene body in P56 mouse forebrain. AD-dependent expressed SNVs were concentrated in the intron region of the *PLCγ1* gene, which indicates enrichment of histone acetylation. Blue vertical bars indicate the locations of the AD-dependent SNV regions of the *PLCγ1* gene. Black bold letters indicate the nucleotides for WT, red bold letters indicate the insertion or substitution nucleotides for AD, and gray letters indicate the deleted nucleotides for AD.

patterns. Although RNA-seq data are needed to analyze RNA splicing, deep learning with given referenced genomic sequences can be a novel alternative for accurate prediction of splicing. Furthermore, an understanding of RNA splicing mechanisms using deep learning could provide novel ways to develop therapeutics and diagnostic procedures for AD.

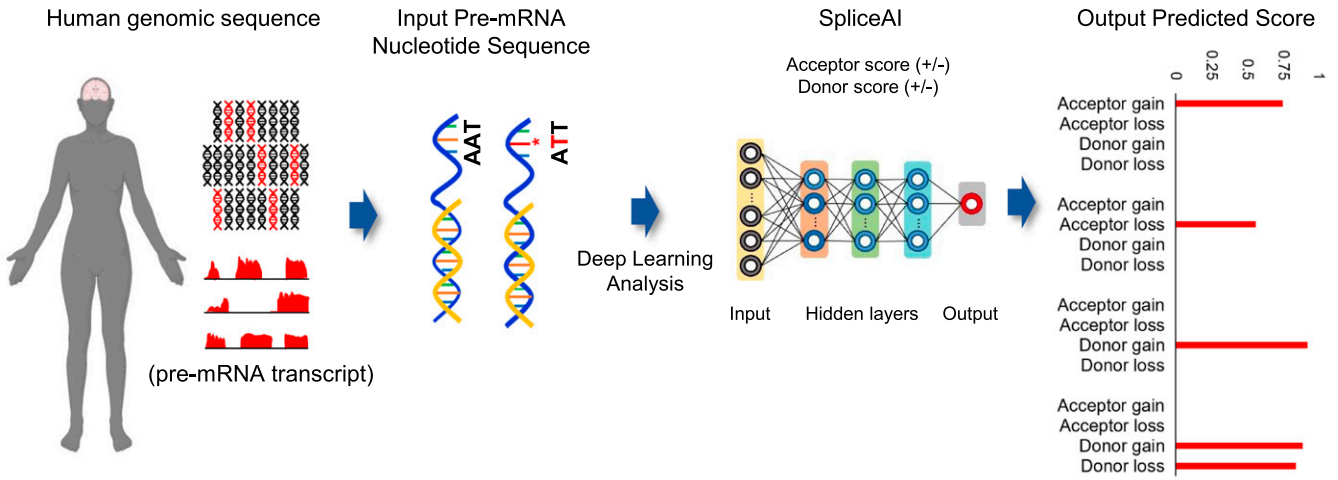
Materials and Methods

Animals. The 5xFAD (6-mo-old) transgenic mice were used as a model of AD. The 5xFAD mice were obtained from the Jackson Laboratory. All animal experiments performed in this study were reviewed and approved by the Institutional Animal Care Use (IACUC) committee at KBRI (Korea Brain Research Institute) (IACUC-20-00018).

Total RNA-seq. Total RNA-seq was performed as previously published (20). Total RNA extraction from mouse cortex was performed using commercial methods based on TRIzol (Invitrogen). Library constructions were performed using TruSeq Stranded Total RNA LT Sample Prep Kit (Human Mouse Rat) according to the manufacturer's instructions (Illumina). Firstly, to make a short fragment of mRNA, we added the fragmentation buffer. The oligo dT-primer was used to synthesize the first-strand complementary DNA (cDNA) to take short fragments as templates. Preparation of synthesis for second-strand cDNA was performed using buffer, 2′-deoxynucleoside 5′-triphosphate (dNTPs, containing 2′-deoxyuridine 5′-triphosphate instead of thymidine 5′-triphosphate), RNaseH, and DNA polymerase I, respectively. Double-stranded cDNA was purified with the QIAQuick PCR extraction kit (Qiagen), and cDNA was eluted by elution buffer. Following the synthesis of the second strand, end repair, addition of single A base, and adaptor ligation, cDNAs were connected with sequencing adaptors. The library concentration was measured by real-time PCR, and the Agilent 2100 Bioanalyzer was used for profiling the distribution of insert size. Library constructions were sequenced with the Illumina HiSeq 4000 based on the manufacturer's instructions (Illumina) and were sequenced for 100 cycles. The HiSeq Control Software HCS (HiSeq Control Software) (v3.3) with RTA (Real-Time Analysis) (v2.7.3) was used to provide the management and execution of the HiSeq 4000 experiment runs.

Sequencing Data Analysis. Total RNA-seq data analysis was performed as previously published (20). Images generated by HiSeq 4000 were converted into nucleotide sequences by base calling and stored in FASTQ format utilizing Illumina package bcl2faseq (v2.16.0.10). To filter the dirty reads, which contain adaptors, unknown or low Phred quality-scored bases were obtained from raw reads, and clean reads were generated. Clean reads were mapped to reference UCSC hg19 genome and gene sequences using Tophat2 (v2.1.0). No more than five bases of mismatched reads were allowed in the alignment. To annotate gene expression, the read values of each gene were calculated as fragments per kb per million using the Cufflinks package (v2.2.1). Fold-change analysis of differentially expressed genes and hierarchical clustering analysis for expression pattern were performed using Agilent GeneSpring (v7.3). Functional enrichment analysis was done using the Gene Ontology functional classification system (geneontology.org/) and DAVID website (https://david.ncifcrf.gov/). Raw reads and data are accessible in the Gene Expression Omnibus (GEO) (accession nos. GSE 151270 and GSE 147792).

SNV Analysis. We performed SNV calls using SAMtools (v.1.3) and bcftools (v.1.3) with bam files from two samples (WT and 5xFAD cortex). Each bam file was indexed with SAMtools, and SAMtools mpileup was performed with UCSC mm10 genome as a reference. To obtain variant calling format (VCF) files for specific target genes, option –r was used to set the position during mpileup. To perform variant calling for five *PLC* genes, five specific regions were set (in detail, chr2:160,731,310 to chr2:160,761,923 for *PLCγ1*, chr2:134,786,164 to chr2:134,988,192 for *PLCβ1*, chr2:118,728,319 to chr2:118,709,202 for *PLCβ2*, chr2:135,741,830 to chr:136,013,068 for *PLCβ4*, and chr19:6,969,643 to chr:6,954,299 for *PLCβ3*), and each of the five VCF files were generated and merged for further analysis. After mpileup, VCF files were generated with vcfutils.pl varFilter using option –D100. The functional annotation of each variant was performed with ANNOVAR software.



Human *PLCγ1* alteration score				Delta Position / Delta score (over 0.5)				
Chrom	Position	Rerence Sequence	Alteration	Acceptor gain	Acceptor loss	Donor gain	Donor loss	Alteration position
chr20	41,159,758	G	A	0	0	0	0 / 0.88	41,159,758
			C	0	0	0	0 / 0.91	41,159,758
			T	0	0	0	0 / 0.85	41,159,758
chr20	41,162,725	G	A	0	0	0	0 / 0.58	41,162,725
			C	0	0	0	0 / 0.56	41,162,725
			T	0	0	0	0 / 0.51	41,162,725
chr20	41,162,992	G	A	0	0	0	0 / 0.79	41,162,992
			C	0	0	0	0 / 0.83	41,162,992
			T	0	0	0	0 / 0.83	41,162,992
chr20	41,163,275	G	A	0	0	0	0 / 0.61	41,163,275
			C	0	0	0	0 / 0.81	41,163,275
			T	0	0	0	0 / 0.78	41,163,275
chr20	41,163,378	G	A	(+3) / 0.75	0	0	0	41,163,381
			C	(+3) / 0.84	0	0	0	41,163,381
			T	(+3) / 0.82	0	0	0	41,163,381
chr20	41,164,006	T	A	0	0	0	(- 2) / 0.68	41,164,004
			C	0	0	0	(- 2) / 0.67	41,164,004
			G	0	0	0	(- 2) / 0.68	41,164,004
chr20	41,165,101	G	A	0	0	0	0 / 0.51	41,165,101
			C	0	0	0	0 / 0.77	41,165,101
			T	0	0	0	0 / 0.61	41,165,101
chr20	41,165,639	G	A	(+5) / 0.41	0	0	0	41,165,644
			C	(+5) / 0.45	0	0	0	41,165,644
			T	(+5) / 0.52	0	0	0	41,165,644
chr20	41,166,394	A	C	0	0	0	0	None
			G	0	0	0	0	None
			T	0	0	(- 2) / 0.89	0	41,166,392
chr20	41,167,852	G	A	(+3) / 0.59	0	0	0	41,167,855
			C	(+3) / 0.92	0	0	0	41,167,855
			T	(+3) / 0.84	0	0	0	41,167,855
chr20	41,167,929	G	A	0	0	0	0 / 0.58	41,167,929
			C	0	0	0	0 / 0.71	41,167,929
			T	0	0	0	0 / 0.87	41,167,929
chr20	41,172,421	A	C	0	(+2) / 0.56	0	0	41,172,423
			G	0	(+2) / 0.56	0	0	41,172,423
			T	0	(+2) / 0.56	0	0	41,172,423
chr20	41,172,877	G	A	0	0	(+27) / 0.92	0	41,172,904
			C	0	0	(+27) / 0.95	0	41,172,904
			T	0	0	(+27) / 0.92	0	41,172,904
chr20	41,173,534	G	A	0	0	(- 4) / 0.89	0 / 0.84	41,173,530 / 41,173,534
			C	0	0	0	0 / 0.91	41,173,534
			T	0	0	0	0 / 0.89	41,173,534

Fig. 5. Prediction of AD-specific nucleotide alteration sites from a human genomic sequence with a deep neural network. (*Top*) Schematic diagrams of the deep neural network procedure from the human genomic sequence database through the SpliceAI analysis. (*Bottom*) The 14 accurate prediction splicing sites in the human *PLCγ1* gene, with delta scores and position information.

SpliceAI Analysis. Human *PLCγ1* was analyzed by SpliceAI to predict RNA splicing types on the genetic variations (4). SpliceAI, a residual convolutional neural network, predicts the probability of RNA splicing taking place in either donor or acceptor DNA sequences. SpliceAI trained the complex patterns of RNA splicing from DNA sequences of 10 kbps around nucleotides of interest from the GENCODE (Genome ENCyclopedia Of DNA Elements)

V24 databases (GRCh38/hg38), including more than 10,000 genes in chromosomes 2, 4, 6, 8, 10 to 22, X, and Y. We conducted the analyses using the pretrained SpliceAI, which was downloaded at https://github.com/Illumina/SpliceAI. We computed the probability of RNA splicing types (probabilities of donor, acceptor, or neither) for the alternate nucleotides on the starting locus of every exon of *PLCγ1* based on the reference sequences (GRCh38/hg38). Then, the probabilities of the variant being splicing altering on acceptor and donor were computed as

$$\Delta Score\ (acceptor\ gain) = \max(a_{alt} - a_{ref}),$$

$$\Delta Score\ (acceptor\ loss) = \max(a_{ref} - a_{alt}),$$

$$\Delta Score\ (donor\ gain) = \max(d_{alt} - d_{ref}),$$

$$\Delta Score\ (donor\ loss) = max(d_{ref} - d_{alt}),$$

where a_{ref}, a_{alt}, d_{ref}, and d_{alt} denote splice acceptor and splice donor probabilities of the reference and alternate nucleotide, respectively. Each $\Delta Score$ is defined as the maximum of difference between the reference and alternate probabilities on splicing sites. To be specific, the gain score is denoted by the maximum disparity of the alternate from the reference probability, which means the probability of corresponding splicing increases by the gain score. Contrary to the gain score, the loss score demonstrates the maximum difference of reference from alternate probability, inferring the probability of the splicing decreased by the loss score. SpliceAI also provided the $\Delta Position$ of acceptor gain, acceptor loss, donor gain, and donor loss, which indicate the location of splicing changes with the probability relative to the position of interest. The positive delta position is downstream of the variant, whereas negative values are upstream.

cDNA Synthesis. The cDNA synthesis for mature mRNA was performed with the PrimeScript first-strand cDNA synthesis kit (Takara) according to the manufacturer's protocol. To synthesize cDNA, we used 1 μg of total RNA from each sample, and added the reaction mixture containing the Oligo dT Primer (2.5 μM), dNTP mixture (0.5 mM), 5× PrimeScript buffer, RNase Inhibitor (20 U) and PrimeScript RTase (200 U). The synthesized cDNA was heat incubated at 42 °C for 60 min and 95 °C for 5 min, amplified with a thermal cycler dice. RNA templates were reverse transcribed into cDNA, which corresponds to pre-mRNA transcript using the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems) according to the protocol described. Following the reference component suggested, each reaction was composed of 10× RT buffer, 25× dNTP Mix (4 mM), 10× RT Random Primers, MultiScribe Reverse Transcriptase (50 U), RNase inhibitor, and adjusted total volume with nuclease-free water. Then, template RNA sample with equal volume was added to set the total reaction volume. The reverse transcription was performed as suggested optimal thermal cycler condition, heat incubated at 25 °C for 10 min, 37 °C for 120 min, and 85 °C for 5 min, amplified with a thermal cycler dice.

RT-PCR. PCR was performed using the Phusion-HF DNA polymerase kit (NEB). Primers employed were *PLCγ1* Ex27-29 forward, 5′-GATTGGCACAGAACGTGCTT-3′, reverse, 5′-CTCCGTCTTTTGCTTGGTGC-3′; *PLCγ1* Ex27-28 forward, 5′-GATTGGCACAGAACGTGCTT-3′, reverse, 5′-GCAAATGACACAGGGTTCCA-3′; *PLCγ1* Ex28-29 forward, 5′-GCAGATGAACCAGGCCC-3′, reverse, 5′-CTCCGTCTTTTGCTTGGTGC-3′; *PLCγ1* Ex26-30 forward, 5′-CTGAGGGGAAGATGATGGA-3′, reverse, 5′-CAGGCCTTTTACTGGGAAAG-3′; *GAPDH* forward, 5′-CATGGCCTTCCGTGTTCCTA-3′, reverse, 5′-GCGGCACGTCAGATCCA-3′; *PLCγ1* Ex27-32 forward, 5′-GATTGGCACAGAACGTGCTT-3′, reverse 5′-CAAATGGCTGCTGGTATCTG-3′. Each reaction mixture contained 5× phusion HF buffer, dNTP (0.2 mM), primer forward and reverse (0.5 μM), Phusion DNA polymerase (1 U), and template cDNA, and nuclease-free water was added. The amplification was performed according to the recommended condition. PCR products were detected through the ChemiDoc (Bio-Rad) imaging system.

RT-qPCR. Total RNA was isolated from mouse whole blood using the TRIzol method. The cDNA synthesis for RT-qPCR was performed with the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems). Primers for RT-qPCR employed were *PLCγ1* coding forward, 5′-GGTGACCTCAGTCCTTTCAG-3′, reverse, 5′- GAAATCTTCAAATGGCTGCTG-3′. RT-qPCR amplification was performed for 10 min for initial denaturation, followed by 45 cycles of 10 s at 95 °C for denaturation, 30 s at 60 °C for annealing, and 1 min at 72 °C for extension. PCR products were analyzed with the LightCycler 480 (Roche).

Western Blot and Antibodies. Electrophoresis was performed with 8% polyacrylamide gels, and proteins were then electrotransferred to nitrocellulose membranes. Membranes were immunoblotted with mouse monoclonal anti-PLCγ1 antibody (B16-5), generated as described previously (49), and mouse monoclonal anti-PLCβ1 (sc-5291, Santa Cruz), rabbit polyclonal anti-PLCβ3 (sc-403, Santa Cruz), rabbit polyclonal anti-PLCβ4 (sc-404, Santa Cruz), and mouse monoclonal anti-β-actin (GT5512, GeneTex) antibodies were washed in Tween 20/Tris-buffered saline containing 5% skim milk. After incubation with the appropriate peroxidase-conjugated secondary antibody, proteins were detected with an enhanced chemiluminescence system (ECL Western Blotting Detection System, Amersham).

ChIP-Seq Data. H3K27ac binding profiles throughout the mouse forebrain were adapted from the data generated by GSE52386 data. H3K27ac ChIP-seq data are available in GEO (https://www.ncbi.nlm.nih.gov/geo/) (50).

Data Availability. RNA sequencing data have been deposited in the Gene Expression Omnibus database (https://www.ncbi.nlm.nih.gov/geo/, GSE 151270 and GSE 147792).

ACKNOWLEDGMENTS. We thank Dr. S.-W. Lee for assistance with bioinformatics analysis. This work was supported by KBRI Basic research program through KBRI funded by the Ministry of Science and ICT (Grant 20-BR-02-13), and Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (Grants 2019R1F1A1059595 and 2017R1D1A1B03030741). Figures were created with Biorender.com.

1. O. Kelemen *et al.*, Function of alternative splicing. *Gene* **514**, 1–30 (2013).
2. E. T. Wang *et al.*, Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
3. S. J. Sanders, G. B. Schwartz, K. K. Farh, Clinical impact of splicing in neurodevelopmental disorders. *Genome Med.* **12**, 36 (2020).
4. K. Jaganathan *et al.*, Predicting splicing from primary sequence with deep learning. *Cell* **176**, 535–548.e24 (2019).
5. R. Wang, Z. Wang, J. Wang, S. Li, SpliceFinder: Ab initio prediction of splice sites using convolutional neural network. *BMC Bioinformatics* **20** (suppl. 23), 652 (2019).
6. P. F. Sullivan, D. H. Geschwind, Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. *Cell* **177**, 162–183 (2019).
7. C. L. Masters *et al.*, Alzheimer's disease. *Nat. Rev. Dis. Primers* **1**, 15056 (2015).
8. A. Serrano-Pozo, M. P. Frosch, E. Masliah, B. T. Hyman, Neuropathological alterations in Alzheimer disease. *Cold Spring Harb. Perspect. Med.* **1**, a006189 (2011).
9. R. U. Haque, A. I. Levey, Alzheimer's disease: A clinical perspective and future nonhuman primate research opportunities. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 26224–26229 (2019).
10. Y. R. Yang *et al.*, Primary phospholipase C and brain disorders. *Adv. Biol. Regul.* **61**, 80–85 (2016).
11. D. S. Kang *et al.*, Netrin-1/DCC-mediated PLCγ1 activation is required for axon guidance and brain structure development. *EMBO Rep.* **19**, e46250 (2018).
12. L. Magno *et al.*, Alzheimer's disease phospholipase C-gamma-2 (PLCG2) protective variant is a functional hypermorph. *Alzheimers Res. Ther.* **11**, 16 (2019).
13. M. Y. Follo *et al.*, Response of high-risk MDS to azacitidine and lenalidomide is impacted by baseline and acquired mutations in a cluster of three inositide-specific genes. *Leukemia* **33**, 2276–2290 (2019).
14. A. Quinquenel *et al.*; French Innovative Leukemia Organization (FILO) CLL Group, Prevalence of *BTK* and *PLCG2* mutations in a real-life CLL cohort still on ibrutinib after 3 years: A FILO group study. *Blood* **134**, 641–644 (2019).
15. J. P. Vaqué *et al.*, PLCG1 mutations in cutaneous T-cell lymphomas. *Blood* **123**, 2034–2043 (2014).
16. S. Behjati *et al.*, Recurrent PTPRB and PLCG1 mutations in angiosarcoma. *Nat. Genet.* **46**, 376–379 (2014).
17. D. Kim *et al.*, Phospholipase C isozymes selectively couple to specific neurotransmitter receptors. *Nature* **389**, 290–293 (1997).
18. H. J. Jang *et al.*, Phospholipase C-γ1 involved in brain disorders. *Adv. Biol. Regul.* **53**, 51–62 (2013).
19. Y. R. Yang *et al.*, Forebrain-specific ablation of phospholipase Cγ1 causes manic-like behavior. *Mol. Psychiatry* **22**, 1473–1482 (2017).
20. K. H. Lim, J. Y. Joo, Predictive potential of circulating Ube2h mRNA as an E2 ubiquitin-conjugating enzyme for diagnosis or treatment of Alzheimer's disease. *Int. J. Mol. Sci.* **21**, 3398 (2020).
21. R. Weissmann *et al.*, Gene expression profiling in the APP/PS1KI mouse model of familial Alzheimer's disease. *J. Alzheimers Dis.* **50**, 397–409 (2016).
22. E. Castillo *et al.*, Comparative profiling of cortical gene expression in Alzheimer's disease patients and mouse models demonstrates a link between amyloidosis and neuroinflammation. *Sci. Rep.* **7**, 17762 (2017).
23. S. K. Han, V. Mancino, M. I. Simon, Phospholipase Cbeta 3 mediates the scratching response activated by the histamine H1 receptor on C-fiber nociceptive neurons. *Neuron* **52**, 691–703 (2006).
24. J. Kim, J. M. Basak, D. M. Holtzman, The role of apolipoprotein E in Alzheimer's disease. *Neuron* **63**, 287–303 (2009).

25. T. Jonsson *et al*., Variant of TREM2 associated with the risk of Alzheimer's disease. *N. Engl. J. Med.* **368**, 107–116 (2013).
26. J. S. Park *et al*., Brain somatic mutations observed in Alzheimer's disease associated with aging and dysregulation of tau phosphorylation. *Nat. Commun.* **10**, 3090 (2019).
27. Z. Wang, M. Gerstein, M. Snyder, RNA-seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
28. F. E. Baralle, J. Giudice, Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* **18**, 437–451 (2017).
29. M. Irimia, D. Penny, S. W. Roy, Coevolution of genomic intron number and splice sites. *Trends Genet.* **23**, 321–325 (2007).
30. L. Shkreta *et al*., Cancer-associated perturbations in alternative pre-messenger RNA splicing. *Cancer Treat. Res.* **158**, 41–94 (2013).
31. M. Petukh, T. G. Kucukkal, E. Alexov, On human disease-causing amino acid variants: Statistical study of sequence and structural patterns. *Hum. Mutat.* **36**, 524–534 (2015).
32. Y. Lee, D. C. Rio, Mechanisms and regulation of alternative pre-mRNA splicing. *Annu. Rev. Biochem.* **84**, 291–323 (2015).
33. K. Marushige, Activation of chromatin by acetylation of histone side chains. *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3937–3941 (1976).
34. X. Lu, L. Wang, C. Yu, D. Yu, G. Yu, Histone acetylation modifiers in the pathogenesis of Alzheimer's disease. *Front. Cell. Neurosci.* **9**, 226 (2015).
35. G. Biamonti *et al*., Alternative splicing in Alzheimer's disease. *Aging Clin. Exp. Res.*, 10.1007/s40520-019-01360-x (2019).
36. M. Montes, B. L. Sanford, D. F. Comiskey, D. S. Chandler, RNA splicing and disease: Animal models to therapies. *Trends Genet.* **35**, 68–87 (2019).
37. G. Fusco, A. Minelli, Phenotypic plasticity in development and evolution: Facts and concepts. Introduction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 547–556 (2010).
38. Y. Liu, A. Beyer, R. Aebersold, On the dependency of cellular protein levels on mRNA abundance. *Cell* **165**, 535–550 (2016).
39. B. Zhang *et al*.; NCI CPTAC, Proteogenomic characterization of human colon and rectal cancer. *Nature* **513**, 382–387 (2014).
40. A. Battle *et al*., Genomic variation. Impact of regulatory variation from RNA to protein. *Science* **347**, 664–667 (2015).
41. J. Y. Joo, K. Schaukowitch, L. Farbiak, G. Kilaru, T. K. Kim, Stimulus-specific combinatorial functionality of neuronal c-fos enhancers. *Nat. Neurosci.* **19**, 75–83 (2016).
42. K. Schaukowitch *et al*., Enhancer RNA facilitates NELF release from immediate early genes. *Mol. Cell* **56**, 29–42 (2014).
43. M. P. Creyghton *et al*., Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 21931–21936 (2010).
44. A. Rada-Iglesias *et al*., A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
45. D. Hnisz *et al*., Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
46. J. Chapuis *et al*.; GERAD consortium, Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology. *Mol. Psychiatry* **18**, 1225–1234 (2013).
47. J. M. Long, D. M. Holtzman, Alzheimer disease: An update on pathobiology and treatment strategies. *Cell* **179**, 312–339 (2019).
48. D. M. Camacho, K. M. Collins, R. K. Powers, J. C. Costello, J. J. Collins, Next-generation machine learning for biological networks. *Cell* **173**, 1581–1592 (2018).
49. P. G. Suh, S. H. Ryu, W. C. Choi, K. Y. Lee, S. G. Rhee, Monoclonal antibodies to three phospholipase C isozymes from bovine brain. *J. Biol. Chem.* **263**, 14497–14504 (1988).
50. A. S. Nord *et al*., Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* **155**, 1521–1531 (2013).